

Variance when variables are not independent

Consider the experiment of rolling two dice, and let X on each roll be the value showing on the first die, and Y on each roll be the sum of the values showing on both dice. As we have seen before,

$$\begin{aligned} 1 \leq X \leq 6 & \quad E(X) = 3.5 & \quad V(X) = 2.92 \\ 2 \leq Y \leq 12 & \quad E(Y) = 7 & \quad V(Y) = 5.84 \end{aligned}$$

X and Y are not independent: $p(X = 1)$ and $p(Y = 12)$ are both nonzero, but $p(X = 1 \ \& \ Y = 12) = 0 \neq p(X = 1) \cdot p(Y = 12)$.

Consider all the values of Y and $X + Y$ in a table where die 1's value is across the top and die 2's value down the left side (values of X are the column headers):

| $Y, X + Y$ | 1 | 2 | 3 | 4 | 5 | 6 |
|------------|------|-------|-------|--------|--------|--------|
| 1 | 2, 3 | 3, 5 | 4, 7 | 5, 9 | 6, 11 | 7, 13 |
| 2 | 3, 4 | 4, 6 | 5, 8 | 6, 10 | 7, 12 | 8, 14 |
| 3 | 4, 5 | 5, 7 | 6, 9 | 7, 11 | 8, 13 | 9, 15 |
| 4 | 5, 6 | 6, 8 | 7, 10 | 8, 12 | 9, 14 | 10, 16 |
| 5 | 6, 7 | 7, 9 | 8, 11 | 9, 13 | 10, 15 | 11, 17 |
| 6 | 7, 8 | 8, 10 | 9, 12 | 10, 14 | 11, 16 | 12, 18 |

You can see in this that overall, higher values of X tend to go with higher values of Y and lower with lower; this makes the distribution of values of $X + Y$ flatter (more pushed to the highest and lowest values) than it would be if X and Y were independent (say, we had three dice and Y was the sum on dice 2 and 3). In particular:

$$\begin{aligned} p(X + Y = k) &= 1/36 \text{ for } k \in \{3, 4, 17, 18\} \\ p(X + Y = k) &= 1/18 \text{ for } k \in \{5, 6, 15, 16\} \\ p(X + Y = k) &= 1/12 \text{ for } k \in \{7, 8, 9, 10, 11, 12, 13, 14\}. \end{aligned}$$

You can check that $E(X + Y) = 10.5 = E(X) + E(Y)$; we don't need independence for adding expected values. We do need independence for multiplying expected values, which is the issue when we consider variance.

To find $V(X + Y)$ directly, find the expected value of $(X + Y)^2$:

$$\begin{aligned} E((X + Y)^2) &= \frac{9 + 16 + 289 + 324}{36} + \frac{25 + 36 + 225 + 256}{18} \\ &\quad + \frac{49 + 64 + 81 + 100 + 121 + 144 + 169 + 196}{12} \\ &= \frac{638}{36} + \frac{542}{18} + \frac{924}{12} = 124.83 \end{aligned}$$

$$V(X + Y) = E((X + Y)^2) - E(X + Y)^2 = 124.83 - 110.25 = 14.58$$

On the other hand, $V(X) + V(Y) = 2.92 + 5.84 = 8.76$, which is smaller, as expected.